

NOVEMBER 2011

WATCHING TEACHERS WORK

Using Observation Tools to Promote Effective Teaching in the Early Years and Early Grades

LISA GUERNSEY AND SUSAN OCHSHORN



EDUCATION POLICY PROGRAM

Early Education Initiative

EarlyEd.NewAmerica.net

NEW AMERICA FOUNDATION

This report is funded through generous grants from the Foundation for Child Development, the W. Clement and Jessie V. Stone Foundation, and the A.L. Mailman Family Foundation.

The New America Foundation would like to thank the following individuals for sharing their time, expertise, and insights for this paper: Holly Blum, Allison Boothe, Janet Burke, Maura Burke, Brenda Cubero, Charlotte Danielson, Laura Ensler, Karen Freel, Barbara Gebhard, Linda Gillespie, Kim Guieb-Kang, Morgan Green, Dana Goldstein, Bridget Hamre, Phyllis Ida, Elizabeth Isakson, Kristie Kauerz, Rebecca Klein, Lourdes Lambert, Tammy Mann, Chris Maxwell, Jack McCarthy, Sara Neville-Morgan, Susan Neuman, Cheryl Ohlson, Bob Pianta, Bentley Ponder, Toni Porter, Sharon Ritchie, Schellee Rocher, Jason Sachs, Eva Marie Shivers, Kathryn Tout, Alicia Tuesta, Christina Weiland, Gerrit Westervelt, Tanya Wright, Virginia Vitrello, Karen Yarbrough, and Martha Zaslow.



© 2011 New America Foundation

This report carries a Creative Commons license, which permits non-commercial re-use of New America content when proper attribution is provided. This means you are free to copy, display and distribute New America's work, or include our content in derivative works, under the following conditions:

- **Attribution.** You must clearly attribute the work to the New America Foundation, and provide a link back to www.Newamerica.net.
- **Noncommercial.** You may not use this work for commercial purposes without explicit prior permission from New America.
- **Share Alike.** If you alter, transform, or build upon this work, you may distribute the resulting work only under a license identical to this one.

For the full legal code of this Creative Commons license, please visit www.creativecommons.org. If you have any questions about citing or re-using New America content, please contact us.

Contents

Executive Summary	1
Identifying Effective Teaching: Observation Tools in Early Education.....	4
Table 1: A Sampling of Tools	6
Table 2: Tools by Age Range.....	6
Promoting Effective Teaching: Professional Development	7
For Educators in Settings for Children from Birth to Age Three.....	7
For Teachers of Three- and Four-Year Olds	8
To Advance PreK-3rd Reforms.....	9
Rewarding Effective Teaching: Formal Evaluation.....	12
For Educators in Settings for Children from Birth to Age Three.....	12
For Teachers of Three- and Four-Year Olds	13
To Advance PreK-3rd Reforms.....	14
The Potential to Enhance Teachers' Effectiveness Throughout PreK-12 Education.....	15
The Goal: Integrating Professional Development and Formal Evaluation	16
Conclusion	17
Recommendations	17
Interviews Conducted.....	20
Notes.....	21

With new research highlighting the connection between children's struggles in the early grades and our nation's high dropout rates,¹ efforts to elevate the teaching of young children should be a high priority. In their formative years, from birth up through third grade, children need caregivers and teachers who can engage them with new concepts and content, attend to skills that need further development, and spark their desire to learn. Yet often these activities take place out of sight, witnessed only by principals or early childhood directors taking stock of a teacher's skills, sometimes based on no more than a few jots about what they see from the doorway.

Objective measurements of teaching and classroom quality are rarely part of larger discussions of public education or teacher effectiveness.² Observation is often sidelined in K-12 debates, and education programs before kindergarten have only recently started to measure what makes for positive interactions between children and teachers.

Imagine how education might change if policies were based on actually watching teachers at work, rewarding good practice and fostering improvements. Studies consistently remind us of what children could achieve if they attended high-quality early learning programs and received high-quality instruction in their early grades of school.³ But the reality is that too many children are experiencing interactions with caregivers and teachers that are inconsistent from year to year and sometimes quite poor.⁴ For children in subsidized child care centers, where staff training is often inadequate, rich learning experiences are a distant second to safety and snack time. Nor are state-funded pre-K programs hitting the mark. A study of programs in 11 states revealed that the average pre-K classroom was not offering children an experience that could be labeled as "good" and only 8 percent of classrooms met criteria that could be described as "good to excellent."⁵

Once children enter elementary school, levels of instruction are not much better. A national evaluation of more than 1,000 elementary school classrooms showed that only 7 percent of children experience consistently good interactions with their teachers, including instructional and emotional support, throughout their elementary

school years.⁶ One recent study of second- and third-grade classrooms in Baltimore highlighted the declining levels of interaction and engagement as testing dates draw near.⁷ By middle childhood, many students have become steeped in these stultifying experiences: In one large national study of 5th grade classrooms, for example, researchers found "positive individual interactions" in only one percent of the time-periods they observed,⁸ with students "spending most of their time sitting around, watching the teacher deal with behavior problems, and engaging in boring and rote instructional activities such as completing worksheets and spelling tests."⁹

For children from economically disadvantaged and minority families, such mediocrity deepens already entrenched achievement gaps, some of which have been detected as early as 9 months of age.¹⁰ With the exception of Head Start, which is designed for the poorest of the poor and is funded to serve only half of those who are eligible, disadvantaged children are typically placed in settings that lack language-rich interactions and learning activities designed to activate a young child's mind.¹¹ Worse, by the time these children arrive in kindergarten, they are likely to attend schools with inexperienced teachers.¹² By fourth grade, according to the National Assessment of Educational Progress, only 17 percent of children from low-income families are reading "proficiently," or at grade level. Scores are similar for black and Hispanic students.¹³

Despite these staggering statistics, current policies are disturbingly silent on how to identify good teaching, pro-

move it, and reward it. Public policy has typically emphasized teachers' education levels and credentials rather than objective measures of how well they teach.¹⁴ Teachers' salaries are tied to years of service as opposed to objective measures of their talents, and public school teachers do not typically receive mentoring related to specific situations in their classrooms. In infant-and-toddler settings, as well as some pre-K settings, caregivers are rarely given time to develop their skills, let alone talk with supervisors or mentors about strategies for improving interactions in their classrooms.

Current policies are disturbingly silent on how to identify good teaching, promote it, and reward it.

To change this dynamic, a growing number of policymakers are searching for new approaches. In the world of programs for children up to age 5, many states have developed Quality Rating and Improvement Systems (QRIS) that identify, rate, and enhance the quality of programs based on a wide array of criteria, such as adult-child ratios and how well teachers respond to children's needs. In the K-12 world, states are trying to identify good teaching at the level of the individual teacher. They are building new and controversial evaluation systems based on "multiple measures" of teachers' abilities, including credentials, portfolios of their work, and, increasingly, growth in students' test scores.¹⁵

Observation tools should play a significant role in the development of these evaluation and professional development systems. These tools can allow for measurements that are far less subjective than many of the checklists and rubrics currently used by supervisors as they pop in and out of classrooms, as long as they include two attributes: They need to be *reliable*, meaning they can be trusted to provide consistent measures of quality no matter who is doing the observing. And they should be *validated*, meaning that studies show their measures to be associated with positive impacts on children's learning, helping them to gain skills in language, literacy, math, social interactions, and other domains.

Across the pre-K through third grade spectrum, observa-

tion tools have the potential to encourage much greater alignment and continuity. When used across early education programs (Head Start, pre-K and child care) and up through kindergarten and the early elementary grades, these instruments can help to create a common language for educators to talk about their teaching, fostering a shared vision of high-quality practice and common standards of professionalism. Today's early education system is weakened by discrepancies between standards and measurement tools used for K-12 teachers and those for professionals in child care and pre-K programs. The use of the same observation tools, across pre-kindergarten and K-12 settings, would help to bridge this gap.¹⁶

Of course, observation tools cannot change the state of early education—let alone PreK-12 education—overnight. The use of these tools will require new mindsets and new funding for the development of systems that include trained observers and careful data collection. Nor is it sufficient to reward quality through evaluation systems that stamp teachers or programs as "good" or "bad" without any emphasis on promoting better practices. Professional development and formal evaluations will need to go hand-in-hand, with data from observations bridging the two.

Valid and reliable observation tools can allow for measurements that are far less subjective than many of the checklists and rubrics currently used by supervisors as they pop in and out of classrooms.

But to lift the quality of education for all our nation's students and to narrow the persistent achievement gaps, teacher observation has to move to a more prominent place in education policy. Objective observation measures can stimulate teachers' reflections and discoveries about where and how to make changes. With the help of coaches and colleagues, teachers can customize strategies for improvement. And when used in formal evaluations, objective observation data can lend credibility to assessments of a teacher's ability to spur children to achieve. Getting at the heart of learning—the interactions between students and their teachers—requires watching teachers work.

To make these improvements, the Early Education Initiative at the New America Foundation recommends the following:

The federal government should:

1. Fund large-scale research and implementation projects on teacher observation tools.
2. Encourage the use of valid and reliable teacher observation tools in the development of evaluation and improvement systems.
3. Highlight and reward the use of the tools to promote alignment of professional standards between early childhood settings and public schools.
4. Highlight and reward the use of the tools to integrate formal evaluation with ongoing, effective professional development.
5. Favor strong examples of the use of the tools in grants and appropriations to teacher preparation programs and professional development providers across the PreK-12 spectrum.

States should:

6. Ensure that new designs for teacher-evaluation systems include the use of valid and reliable observational tools.
7. Continue to develop and refine Quality Rating and Improvement Systems (QRIS) using valid and reliable observation tools that focus on adult-child interactions.
8. Establish guidelines for using observation tools in accordance with research-based examples of best practice.
9. Dedicate funds for the development and sustainability of comprehensive assessment and improvement systems that use valid and reliable observation tools.

Local educators and leaders (including school district leaders, principals, and directors of early childhood programs) should:

10. Participate in training on the use of observation tools to gain a greater understanding of the types of interactions and teaching strategies that foster learning.
11. Give teachers and caregivers opportunities to be observed and assessed using valid and reliable tools.

Teacher-preparation programs should:

12. Customize coursework and clinical experiences for prospective teachers to familiarize them with observation tools, classroom strategies that enhance teacher-child interactions, and the importance of reflecting on one's own teaching.
13. Require prospective teachers to be observed and assessed using valid and reliable tools and provide research-based strategies for improvement.

Developers of professional development initiatives should:

14. Use valid and reliable observation tools to customize interventions for teachers.

Researchers should:

15. Study the validity and reliability of existing observation tools in multiple settings and with children of varying backgrounds, including English language learners.
16. Continue to develop observation tools based on the latest findings in science.
17. Expand research on the tools to include teacher aides, assistant teachers, directors, principals, and other administrators in the public schools, as well as professionals in multiple types of child care settings. ■

Identifying Effective Teaching: Observation Tools in Early Education

It's a spring morning in Angelique's (not her real name) classroom in Alexandria, VA, where three- and four-year-old children are gathering at her feet in a circle on the rug, coaxed by an assistant teacher. At a table near the center of the room sits a professional observer with an open laptop. The children notice this person at first but soon seem to forget she is there. The observer scans the room, types for a moment, scans again, and then focuses on what Angelique is saying to the children around her. Over the next hour-and-a-half, this observer will watch and write notes as the children break up into small groups and move to various areas of the room—the plant and animal area near the windowsill, the water table near the front door, the reading corner, the “dress-up” area where children can pull out costumes of their choice. The observer will watch the teacher deftly defuse a conflict between two children who want to wear the same princess dress. She will take note of the way the teacher elaborates on children's comments while filling pitchers at the water table. By the end of the observation, dozens of data points from that observation will provide a snapshot of the classroom environment, including the instructional skills and level of attunement of the teacher.

Throughout the country, in Head Start classrooms like Angelique's, federal monitors are conducting triennial views using observational tools. Indeed, this type of measurement is an increasingly common phenomenon in pre-kindergarten classes funded with public dollars. Over the past few decades, more than 50 observation tools have been designed for pre-kindergarten settings, and are being put to use in the early grades of elementary school as well.¹⁷ The tools are essentially rubrics that observers use to assign numeric values to tightly defined teacher behaviors or elements of classroom organization.

The Classroom Assessment Scoring System (CLASS), for example, uses a scale from 1 to 7 and is designed to measure interactions within three domains (emotional climate, classroom organization, and instructional support) along multiple dimensions, including quality of feedback and concept development, among others.¹⁸ A score of 1 under the “quality of feedback” dimension reflects a teacher's inability to provide anything more than perfunctory responses to children's questions about something they are doing in class. A score of 7 means that a teacher is often helping children to reach new levels of understanding by engag-

ing in frequent back-and-forth exchanges that genuinely address a child's questions and curiosity.

Other tools for early childhood settings (including kindergarten, but not often beyond) focus on how rooms are arranged for play and privacy; how many books are available at a child's level, how staff members greet children and parents; the amount of time allotted to individual and group work; and informal use of language. Known generically as environmental rating scales (ERS), they are widely used as measures of healthy, safe, and productive environments for young children. Some of them focus on support for children's growth, development, and general well-being, including attention paid to physical activity and good nutrition.

Studies around the country show that teachers get low scores on their ability to promote higher-order thinking skills, offer quality feedback, and provide models for using language well.

The types of observation tools available today range across settings and age groups, offering a veritable alphabet soup of acronyms. (See Table 1 on p. 6) In addition to the CLASS, for example, there are several others, including but not limited to the Early Childhood Rating Scale (ECERS); the accreditation tool used by the National Association for the Education of Young Children (NAEYC); the Quality Indicators (QI) tool; and the Program Quality Assessment (PQA) for infant/toddler and pre-kindergarten settings. The Family Child Care Environmental Rating Scale (FCCERS) is for family child care and other home-based settings. The Program Assessment Rating Scale (PARS) is for birth-to-three settings, as is the Child/Home Early Language and Literacy Observation (CHELLO), which can also be used in home settings. The Infant/Toddler Environmental Scale (ITERS) is targeted to the birth-to-age-three end of the spectrum. The CLASS, Snapshot and the Early Language and Literacy Classroom Observation (ELLCO) instruments can be used in pre-K and K-3 classrooms.

Some tools are specifically geared toward gathering data on how teachers interact with students in various circum-

stances, and recent research shows how important that specificity can be. Studies using the CLASS have already pinpointed at least one element of teaching that needs major improvement in most classrooms for three- and four-year-olds around the country: instructional support. Studies conducted in thousands of classrooms show that teachers score in the moderate-to-high range in organizing classroom time and providing emotional support to children, but get low scores on their ability to promote higher-order thinking skills, offer quality feedback, and provide models for using language well¹⁹—all of which are necessary to support language development and ensure that children are building a foundation for school success.

How well these tools identify good teaching with English language learners is a question in need of further exploration.

The availability of tools that can reliably capture information on effectiveness of child-teacher interactions represents a huge advance for early childhood programs, and education in general. But studies of their validity—their association with good outcomes for young children—are still emerging. While a number of studies offer promising evidence,²⁰ some researchers have pointed out how difficult it can be to draw a straight line from a teacher's scores during observations to children's proficiency in math, reading and social emotional development.²¹

Observation tools alone are not the end-all-be-all. They can promote good teaching only when used appropriately by trained observers, integrated well into ongoing professional development, and incorporated into existing education structures by policymakers who understand both their promise and limitations. Some tools may not work well in

family- and home-based care settings. Good teaching may look different when practiced with smaller groups of children who vary widely by age—in home-based settings, for example. And some tools may not provide enough information on best practice in specific content areas, such as math, literacy or science, where different approaches to teaching may be required. Strong teaching of, say, first- or second-grade math may require a deep understanding of numeracy, yet current rubrics may not be precise enough to capture that depth of knowledge.

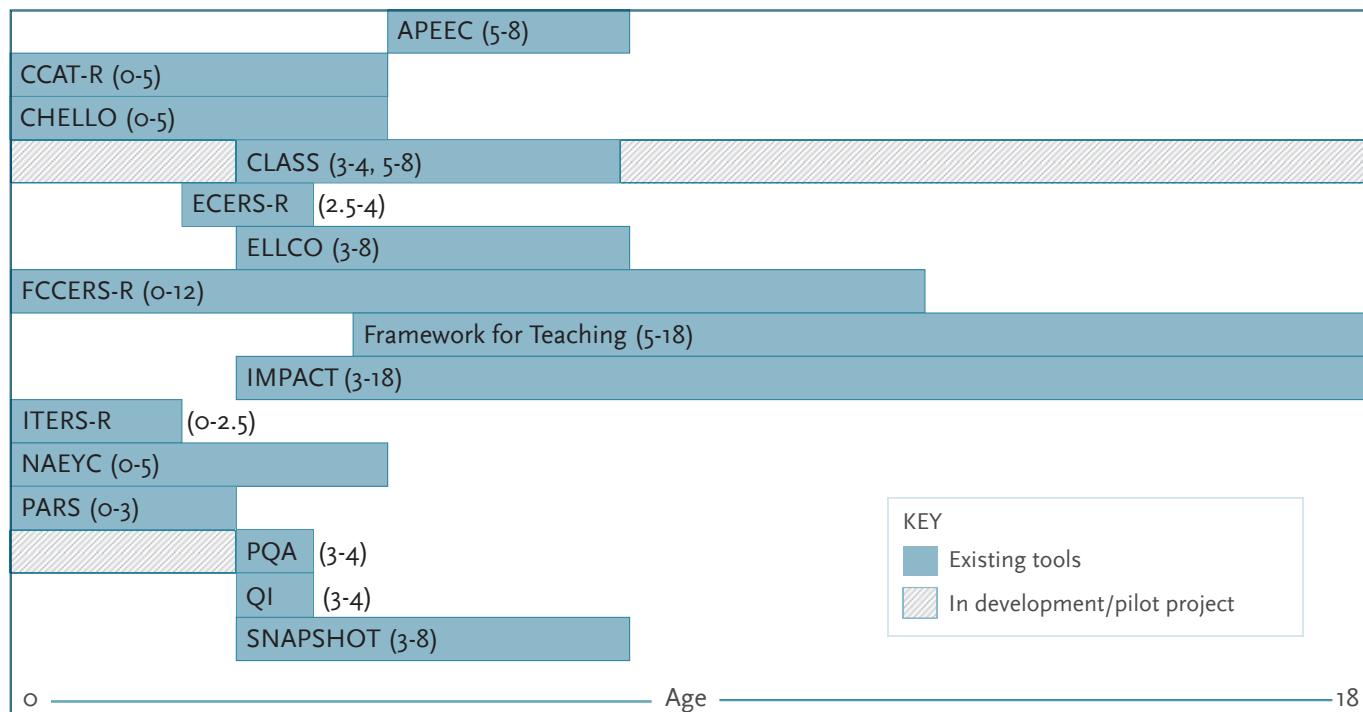
Moreover, over 20 percent of children in the U.S. today speak a language other than English at home²² and their numbers are growing. In some parts of the country, Latino children are already the majority of the school-aged population. How well these tools identify good teaching with English language learners is a question in need of further exploration. Ensuring the quality of their educational experiences will be critical for the workforce of the future.

Lastly, policymakers and educators need to recognize that building observation-based systems of professional development and evaluation—as is the case with any assessment system or teacher-training program in education—will require a significant outlay of funds. The expense of simply performing the observations can depend on multiple variables, from the number of observers that require training to the extent of travel required for visits to geographically disparate programs and classrooms to the cost of writing and filing reports that make sense of the data. In QRIS systems, observations can cost several hundred dollars to a few thousand dollars per classroom.²³ The cost of professional development varies widely as well, depending on how often coaches are in contact with teachers, what level of guidance they provide related to the observation measures, and whether they offer help through video chats and online connections instead of in person. ♦

Table 1: A Sampling of Tools

Instrument	Publisher/Source	Classroom	Home
Assessment of Practices in Early Elementary Classrooms (APEEC)	TC Press	✓	
Child Care Assessment Tool for Relatives (CCAT-R)	Bank Street College		✓
Child/Home Early Language and Literacy Observation (CHELLO)	Brookes		✓
Classroom Assessment Scoring System (CLASS)	Brookes	✓	
Early Childhood Environment Rating Scale—Revised (ECERS-R)	TC Press	✓	
Early Language and Literacy Classroom Observation Toolkit (ELLCO)	Brookes	✓	
Family Child Care Environment Rating Scale—Revised (FCCERS-R)	TC Press		✓
Framework for Teaching	ASCD	✓	
IMPACT Observations	District of Columbia Public Schools	✓	
Infant/Toddler Environment Rating Scale—Revised (ITERS-R)	TC Press	✓	
NAEYC Accreditation Observation (NAEYC)	National Association for the Education of Young Children	✓	
Program Assessment Rating Scale (PARS)	West Ed	✓	✓
Program Quality Assessment (PQA)	HighScope Educational Research Foundation	✓	
Quality Indicators (QI)	AppleTree Institute for Education Innovation	✓	
Snapshot	FirstSchool	✓	

Table 2: Tools by Age Range



Source for charts: New America Foundation reporting; *Quality Measurement in Early Childhood Settings* (Brookes, 2011); and *Quality in Early Childhood Care and Education Settings: Compendium of Measures, Second Edition* (OPRE, 2010).

Promoting Effective Teaching: Professional Development

Early evidence hints that with good coaching and other individualized forms of professional development, observation tools have the potential to be powerful levers for improving programs and teachers' practice.²⁴ Teachers who are observed using these tools are more likely to receive assessments in sync with the most current knowledge about effective professional development for early childhood educators. The scores and other results from these assessments serve as a barometer of teachers' strengths and weaknesses. This information can stimulate conversations among practitioners about how to improve and guide their work and the work of professional development specialists. Ultimately, this data can help determine the effectiveness of professional development models in changing the way teachers teach—information that is highly useful to policymakers trying to prioritize investment of limited funds.

For Educators of Children from Birth to Age Three

Janet Burke, a rater for the Virginia's Quality Rating and Improvement System, once observed a teacher working with 2-and-a-half-year-olds who were standing at a sink in their child care center. She noticed a boy pushing a plastic toy down into the water and watching it pop back up. His teacher offered more toys and encouraged the boy and his peers to experiment. "If this one sinks," she asked the toddlers, "what do you think would happen if we put the two together?" Burke was impressed. "You could just see this little boy's brain moving," she said. For Burke and early educators who work with infants and toddlers, scenes like this are evidence of the importance of good teaching even at very young ages.

Yet to date, only a few observation tools have been specifically designed for the earliest end of the education spectrum. Of those that exist, very few sufficiently capture the interactions of caregivers and children,²⁵ and few are used to train caregivers on how to improve those interactions. The CLASS instruments are an exception, with an infant tool under development and the recent release of a toddler tool. The professional development program called My Teaching Partner (MTP) is designed to work with the CLASS across different age groups. In MTP, teachers are assigned to coaches who observe them, via video or in person, and coach them—online, in person, or on the phone—on strategies to sharpen the quality of their teaching skills. In California, a statewide early childhood initia-

tive plans to officially roll out the Toddler CLASS with MTP in the 2012-2013 academic year. In the meantime, the state is working to reach thousands of teachers by introducing them this fall to the tool and an online video library that offers in-the-classroom views of good teaching.

Another tool for infant-and-toddler settings is the Program Assessment Rating Scale (PARS), scheduled for release in the spring of 2012.²⁶ One of the PARS subscales—"quality of caregivers' interaction with infants"—is solely rated by observation.²⁷ Independent assessors spend three or four hours in a classroom observing how care is provided during the daily routines, including greeting and drop-off, feeding and mealtime, diaper changing and toileting, and play. The PARS is designed to accompany the Program for Infant/Toddler Care, a professional development model designed in the late 1980s to bring multimedia training materials to center- and home-based childcare providers.²⁸ Program participants are assessed with the PARS before and after going through the program.²⁹ A study using PARS in California revealed significant improvements in overall quality, with the most consistent positive change in the quality of teachers' and caregivers' interactions with infants and toddlers.³⁰ In South Carolina, professional development specialists use the PARS during their visits to infant-toddler programs. Results from the PARS are passed to the program directors so they can continue the conversations when the specialists depart. The teachers in these programs also observe each other using the PARS as their tool.

Observation tools have the potential to change the mindset of people who work with infants and toddlers. Amy Dombro, co-author of *Powerful Interactions: How to Connect with Children to Extend Their Learning* and an expert on early childhood education, argues that many practitioners don't see themselves as decision makers. And yet to be a good teacher, "you need to be able to put aspects of your practice into words, to be able to think about it," she said, before you can make decisions about how to adjust your interactions with a child. The intersection of observational assessment and professional development can bridge this gap, providing a rich opportunity for teachers of infants and toddlers to gain a clearer conception of how their interactions can make a difference. It is here, in the earliest years of learning, where children's development is most dynamic and elastic, where the cognitive, social-emotional, and physical are inextricably linked, and where interactions with adults can yield rich results.

In Louisiana: Observing Teachers' Sensitivity to Children's Mental Health

When child care professionals do not understand how young children learn and develop, troubling stories emerge. At a ZERO TO THREE conference last year an early childhood expert recounted tales of very young children—some only 18-months old—being “kicked out of child care” reportedly because their actions didn’t conform to a teacher’s perception of good behavior.³¹ These kinds of episodes, also reported in state-funded pre-K programs,³² can seriously disturb a child’s ability to learn, as shown by myriad studies on the connections between young children’s cognitive and social-emotional development.

In Louisiana, to ensure that teachers in child care centers are equipped to handle challenging situations with young children, the state developed an initiative to support mental health that uses data from observation tools. The initiative is now part of the state’s Quality Rating and Improvement System. Programs in the state’s system are rated using a social-emotional subscale created by Geoffrey Nagle, director of Tulane’s Institute of Infant and Early Childhood Mental Health, and Angela Keyes, assistant professor of psychiatry and neurology. Designed in consultation with the creators of the Environmental Rating Scales, this subscale includes domains finely focused on the interactions between caregivers and children and their connection to the environment, such as a caregiver’s level of affection and tone of voice and the extent to which a caregiver seems to be emotionally connected to a child.

Scores on this new subscale can help to guide the services of Infant Mental Health specialists. These specialists conduct 12 full-day visits every other week, over a six-month period. They use the CLASS tool to record what is happening in the classrooms. The specialists also have individual meetings with teachers and families to make them aware of strategies that support positive behaviors and other approaches that have been shown to be successful with young children.³³ “The system is about the social-emotional aspect of children’s development,” Nagle said.

An evaluation of Louisiana’s initiative, to be released in the journal *Early Education and Development*, found that caregivers’ scores on the CLASS showed improvement in positive climate, behavior management, productivity, teacher sensitivity, regard for student perspective, and instructional learning. The most significant changes occurred in the dimensions of positive climate, regard for student perspective, productivity, and instructional learning.³⁴ The state paid Tulane \$125,000 a year to conduct the evaluation.

For Teachers of Three- and Four-Year-Olds

In contrast with the infant-toddler realm, programs that enroll children in their two years prior to kindergarten are far more likely to use observation-based assessment of teachers and classrooms. At least 26 states³⁵ have established or are developing Quality Rating and Improvement Systems (QRIS) that require early childhood programs to undergo observation by trained professionals every one or two years. The rollout of QRIS has led, over the past several years, to trained observers entering thousands of state-funded pre-K and Head Start³⁶ programs around the country armed with these instruments. But even in states that do not have statewide rating systems or are in the process of developing them, the use of observation tools is increas-

ingly common. In Georgia, for example, the Department of Early Care and Learning is sending trained observers to administer the CLASS in all 4,200 classrooms that are part of the state-funded pre-K program.³⁷

The results from these myriad observations are making their way into professional development programs, and evidence is emerging of some promising effects on teaching. A recent study of 400 pre-K teachers in eight urban centers around the country showed that those using MTP improved the teachers’ instructional abilities significantly.³⁸ The coaching within MTP led to improvements in developing concepts with young children, providing more detailed feedback about their achievements, and modeling the use

of language. Teachers gained in their ability to provide emotional support. It should be noted, however, that the study also turned up a small negative association between MTP coaching and the teachers' focus on literacy.

Another model is Michigan's Great Start program, where early educators across the state are taking courses and receiving on-site, one-on-one coaching that is informed, in part, by the use of two observation tools—the ELLCO for center-based programs and the CHELLO for child care provided in the homes of licensed providers. A study of 291 Great Start programs in four cities found no change in teacher knowledge, but significant improvement in teacher practice. Coaching seemed to make the critical difference, a finding confirmed by a replication of this study on different samples in six cities in Michigan.³⁹

There's always room for improvement, however, as shown in a different Great Start study that was designed to tease out the effects of coaching alone on teacher practice. This study looked at the online logs of coaches in 58 early childhood centers. The notes in those logs provided hints as to whether coaches were able to move teachers forward. After 10 coaching sessions in as many weeks, evaluations of the teachers using the ELLCO and the CHELLO revealed that the teachers had made some strides in improving the physical environment for reading and writing. But according to the logs, the teachers made no progress on improving their language and literacy instructional strategies. Additional investigation is needed into how to improve pedagogy.⁴⁰

Observation-based professional development has also become an engine of change at the AppleTree Institute for Education Innovation in Washington, D.C., which runs a network of seven charter preschools. AppleTree has created a professional development system called the Quality Indicators (QI), which uses a home-grown observation tool that was designed to align with features of the CLASS, ELLCO, and the Sheltered Instruction Observation Protocol (which assesses how well teachers are able to engage with children who speak languages other than English). Each AppleTree campus is using Quality Indicators to provide feedback to teachers and center directors, and with the help of a \$5 million federal grant and other funding, AppleTree is building the QI into a professional development system to be used by preschools around the country. So far, data confirms the tool's validity: In a recent analysis of students' performance on several assessments of language and liter-

acy skills, the teachers with the highest QI scores were the ones whose students had the highest achievement scores.⁴¹

In addition to testing whether observation-based professional development changes teaching, researchers are also examining how much it could lead to positive changes in children's achievement. Last year, researchers Amanda Williford and Andrew Mashburn, at the Center for Advanced Study of Teaching and Learning at the University of Virginia, examined the impact of mentoring and observation tools on children's outcomes. They examined data from nearly 50 childcare and pre-K centers in Hampton Roads, Va., half of which were using mentors to help preschool teachers learn from their scores in the state's rating system, which uses both the CLASS and ECERS observation tools. The study collected data from multiple teachers to arrive at scores for programs as a whole. Compared to programs that did not use mentors, Williford and Mashburn found that these programs scored higher on both the CLASS and ECERS after one or two years of intensive mentoring. The average score on the CLASS's instructional support indicator rose from around 2.8 to 3.8 for centers that had two years of mentoring—still not close to the high range (7 is the highest possible score) but better than centers that received no assistance.⁴² More importantly, those centers graduated children who showed significant gains in language development and social-emotional skills compared to children in programs that did not use mentors.

Early childhood centers that used observation tools plus mentoring graduated children who showed significant gains in language development and social-emotional skills compared to children in programs that did not use mentors.

Advancing PreK-3rd Reforms

Observation-based professional development can also be embedded in larger and more comprehensive models for reforming public education. Standard measures of teaching quality are important for any reform effort that extends beyond one grade level. They allow administrators and teachers to compare teaching quality across grades vertically—from kindergarten to first grade, for example—

while also providing common terminology and goals, horizontally, for teachers within those grades. In fact, the use of reliable observation instruments holds promise for enhancing the effectiveness of instruction along the entire spectrum of PreK-12 education.

A growing number of initiatives around the country are using observation-based assessments to jump-start comprehensive changes across pre-kindergarten and the early grades. Examples can be found in schools in large cities

such as Chicago and Boston, as well as peppered across states, including Hawaii, Michigan, and North Carolina. In each case, professional development starts with trained professionals conducting an observation of teachers in the classroom. Teachers receive the results of those observations during consultations with coaches and other professional development specialists, some of whom are assigned to specific teachers and charged with advising those teachers on new instructional approaches to try out in their classrooms.

FirstSchool: A PreK-3rd Approach to Improving Instruction School-Wide

Instilling a culture of high-quality instruction for pre-K through third grade requires intensive collaborations among teachers, principals, and parents. The FirstSchool program based at the Frank Porter Graham (FPG) Child Development Institute at the University of North Carolina was designed to facilitate those collaborations by providing professional development that includes one-on-one coaching for teachers as well as assistance to principals. Over the past two years, eight schools in North Carolina and Michigan have signed up to be “FirstSchool” schools.

To assess teaching, FirstSchool uses a tool called the FirstSchool Snapshot, originally known as the Emerging Academic Snapshot.⁴³ The tool captures information on how teachers use classroom time, with trained observers documenting the activities of four children, minute by minute, throughout an entire school day, coding for both the content (phonics, whole language, math, etc.) and the type of activity (small-group, whole group, outside play, etc). Observers also record how teachers interact with children—the extent to which they ask open-ended questions and encourage children to elaborate or the extent to which they respond harshly or irritably to children individually or in a group.

Trained observers do Snapshot assessments of every classroom in pre-K through the third grade at the beginning and end of the school year. The schools receive visits from assigned coaches and facilitators on a regular basis to talk with teachers, principals, and other staff members about the results of the Snapshot, provide ideas for more effective activities and classroom organization, and engage families in conversations about how their school could improve. FirstSchool also provides data to the schools from the CLASS, teacher and family surveys, and parent focus groups and helps schools examine the results to help determine next steps in meeting their goals. The data triggers discussions among teachers and principals, facilitated by coaches, about how to change what is happening in the classroom. In many cases, for example, the data has shown that students are getting few opportunities to demonstrate their learning and understanding verbally, graphically, or pictorially, or to engage in higher order thinking skills. Instead, large portions of the day are devoted to didactic instruction in whole group settings and transitions.

Snapshot results that might show a change in teachers’ practice are not yet available, but already teachers are exchanging their demands for “hands in your lap, and silence” for more flexibility in where and how children work, according to Sharon Ritchie, director of the FirstSchool project. Children are also given more opportunities for extended conversations, both socially and academically, which offer increased opportunities to use expressive language, build their vocabularies, and collaborate with peers. “We work to help educators think beyond traditional practices,” Ritchie said.

In the Boston Public Schools, for example, efforts are underway to improve the quality of instruction in pre-K and kindergarten classrooms. Jason Sachs, Boston's director of early childhood education, has decided to put these classrooms under multiple microscopes to examine how teachers are doing and how classrooms and daily routines are set up. Every two years, classrooms are evaluated using observations gleaned from the CLASS, ELLCO, and ECERS, as well as through accreditation visits by the National Association for the Education of Young Children.⁴⁴ The results tell Sachs what level of professional development is needed in each school.

In the Boston Public Schools, results from observational assessments are stimulating conversations among teachers and principals about how to improve.

Results also stimulate conversations among teachers and principals about how to improve. During an annual meeting of pre-K and kindergarten teachers, for example, Sachs displayed data showing that only one-third of teachers were scoring in the middle-to-high range in the CLASS's "instructional support" category. Why, he asked, was this happening? "We had a comprehensive discussion about the structure of school, the lack of time, the inability to sustain conversations with children," Sachs said. "Some of the issues have to do with the way schools are set up and some of it has to do with the ways our teachers are thinking about children." Sachs added that he has found NAEYC's standards for accreditation, and the professional development required to meet them, to be the most effective method for promoting better teaching, particularly in kindergarten.

In Chicago, reforms to improve instruction have extended beyond pre-K and kindergarten, into the primary grades. Here, six schools are part of the Erikson Institute's New Schools Project, which adopts schools within the city whose principals have expressed an interest in revamping teaching practices. According to director Chris Maxwell, they are striving to elevate their teaching so that it is both more "developmentally informed" (or attuned to children's

abilities at different stages of development) and intellectually challenging. So far, some teachers in a subset of schools have been observed using the CLASS and coaches have shared the results during professional development sessions. Leaders in some New Schools' schools are also looking at the Framework for Teaching, an observation tool used in a growing number of K-12 settings and potentially employable in pre-K, as an instrument to be combined with the CLASS. Maxwell said she continues to look for the most cost-effective, research-based and standardized way to observe and share evidence with teachers about their work, especially for the teaching of specific content areas, such as math and reading. "We are desperately in search of observation tools to use," she said.

In the next few years, educators expect to learn more about observation-based professional development from several PreK-3rd pilot projects. Administrators are already discovering that it can take years to prepare for school- and district-wide use. At the Farrington Complex in Honolulu—a site of nine elementary schools and 23 pre-K classrooms—officials are rolling out the use of the CLASS one grade-level at a time and are taking care to make sure that the teachers, and principals where possible, are trained in how the observation tools work before assessments are made. That way, said Kim Guieb-Kang, coordinator of the Farrington project, "teachers know what they are being observed for." At some sites in Hawaii, teachers in pre-K and the early grades can get credits that lead to a salary boost for taking courses on the use of the CLASS. The courses require them to videotape their teaching, "score" themselves using the tool's rubrics, share their videos with their peers to compare scores, and choose an area to improve. Teachers call the experience "eye-opening." As one kindergarten teacher wrote in her portfolio after taking the course: "I am now more aware of how I teach and how I deliver the message across students."⁴⁵

The next challenge is applying knowledge to practice. Coaching and other individualized professional development during the school year, accompanied by observations, are an integral part of the process. To ensure sustainability, school officials will have to regularly recruit and train observers, and test their reliability. In Hawaii, schools are tapping their consultant-type teachers, sometimes known as "resource teachers," to become CLASS observers and provide advice based on the results they collect. ♦

Rewarding Effective Teaching: Formal Evaluation

It's one thing to talk about using observation-based tools to improve teaching. It's quite another to rely on them for evaluations of teachers or programs that have "high stakes" attached. Educators are unnerved by the use of data to determine rewards and penalties, promotions and demotions—especially in child care, pre-kindergarten and kindergarten programs, which struggle to find sufficient funding to meet standards and pay for professional development. And yet policymakers need methods for prioritizing funding in publicly funded programs and supervisors need objective ways to identify effective teaching. Setting those priorities means establishing incentives for strong programs and high-quality teaching. Research has shown how much a child's life trajectory can be altered by effective, well-designed early learning programs and good early elementary classrooms.⁴⁶ The scarcity of public funding for early education demands that dollars be channeled to programs that ensure teachers are interacting with children in ways that give them the skills to thrive academically and socially during their school years and beyond.

The scarcity of public funding for early education demands that dollars be channeled to programs that ensure teachers are interacting with children in ways that give them the skills to thrive academically and socially during their school years and beyond.

Two different approaches to evaluation exist within the span of early education offerings for children from birth through 3rd grade. Within programs that serve children from birth to age 5, observation-based assessments of individual teachers are rarely conducted for personnel purposes. Instead, the results from formal observations of teachers are increasingly used to rate the programs where those teachers work, including pre-K and child care centers. Within kindergarten and the early grades, where teachers are not typically part of "programs" that can be evaluated, principals do conduct observations of teachers to make personnel decisions. The use of valid and reliable tools, however, is not common, and observations are rarely

conducted by independent professionals who have been tested for their reliability as consistent raters.

Given these differences, creating a seamless system of evaluation in early education will be difficult, especially across the span of pre-K to third grade. The current state of flux in evaluating public schools and teachers adds to the challenge. As states develop new teacher-evaluation systems for their public schools, they need to grapple with how to do a better job of evaluating teachers in the earliest grades, and possibly in pre-K classrooms based in public schools. Meanwhile, at the beginning of the spectrum—in programs for children from birth to age 5—observation-based assessments of full programs have opened a Pandora's box of new questions for policymakers about where to invest and how to ensure that low ratings do not stymie program efforts at improvement. As a national early childhood task force warned in 2007, "accountability requires great care."⁴⁷ A recent report from the Administration for Children and Family's Office of Policy, Research and Evaluation put it this way: "If programs and teachers do not have confidence that they are being assessed fairly and consistently, the whole system will be undermined."⁴⁸

In Settings for Children from Birth to Age Three

Extra care may be required in using observation instruments to evaluate birth-to-three settings where observation for professional development, much less evaluation, is not common. The majority of U.S. children under five with employed mothers are in home-based settings, where a caregiver may take care of multiple children at one time while also caring for his or her own children, or in less structured care provided by family, friends, and neighbors. Nearly half of all infants and almost 40 percent of toddlers are in a relative's care at least once a week, with the majority of infants cared for by grandmothers, and only a third of infants and toddlers cared for by nonrelatives. Training and education levels among birth-to-three educators, especially the home-based, tend to be among the lowest in the field.⁴⁹

Today, evaluation of birth-to-three settings, if it happens at all, takes the form of self-assessment (required of programs funded through the federal Child Care Development Block Grant (CCDBG)), licensing inspections (which focus on health, safety, adult-child ratios, and other environmental elements, not teaching quality) or intensive third-party observation as part of a state's Quality Rating and Improvement System.

State QRIS may offer the best opportunity for the use of valid and reliable observation tools to evaluate teaching and caregiving in infant and toddler settings. Twenty-one states now require infant-and-toddler programs to be evaluated as part of these systems.⁵⁰ Most of these states use the Infant/Toddler Environmental Rating Scale (ITERS) and the Family Child Care Environmental Scale (FCCERS). But these tools are not enough. Educators worry that “the environmental rating scales don’t capture the teacher-child interactions that get to the heart of quality,” according to Diana Schaack, a researcher at San Francisco State University who evaluated Colorado’s QRIS and has examined the reliability of the ITERS in that high-stakes context.⁵¹ New instruments are emerging to fill the gap. Among them are the Infant and Toddler versions of the CLASS⁵² and the PARS⁵³. Others include the beta version of the Infant-Toddler PQA and the Quality of Caregiver-Child Interactions for Infants and Toddlers (Q-CCI), to be field-tested in 2012. For evaluation of child-caregiver interactions in family-based child care, programs are adopting the Child Care Assessment Tool for Relatives (CCAT-R)

(see box below), and the Quality of Early Childhood Care Settings (QUEST).⁵⁴

The new tools are welcome, but no matter how advanced they become, policymakers shouldn’t see observation-based assessment as a silver bullet. High-stakes evaluations should always take into account multiple indicators. Consider staff turnover, a problem that is particularly acute within infant-and-toddler care. A recent study based on data from Colorado’s QRIS, for example, found significant movement among children and teachers in and out of classrooms, thwarting the kind of consistency and stability critical to the healthy development and early learning of infants and toddlers. This instability would not be captured by observing caregivers one or two times a year.

For Teachers of Three- and Four-Year-Olds

In settings for 3- and 4-year-olds, the use of observation instruments for evaluation purposes is becoming more common. In addition to being assessed under the QRIS systems that have been established in 26 states, pre-K

In Hawaii, Evaluating a Program that Helps Grandparents and Parents Prepare Children for School

An innovative program in Hawaii provides some insight into how formal evaluations can improve learning environments for very young children in family-based settings. The program, Tutu and Me, was created in 2001 as a school-readiness intervention in a state where a disturbingly high 40 percent of kindergartners entered school unprepared. Tutu, which means “grandparent” in Hawaiian, primarily serves children age 3 and under along with their caregivers, who are typically parents and grandparents. Program services, offered over a period of 11 months, include two-hour biweekly sessions in which these caregivers and children interact in a variety of activities, receiving mini-lectures on aspects of child development, access to caregiver resources, and assessments of how well their children are developing.

To determine the effectiveness of their program, officials at Tutu and Me approached the Institute of the Child Care Continuum at Bank Street College of Education, in New York City, to conduct an evaluation.⁵⁵ Working collaboratively, the Hawaiian program and the institute chose to use the Child Care Assessment Tool for Relatives (CCAT-R). The institute provided training on the CCAT-R for program staff, who then integrated its use into ongoing program evaluation. The staff members received training on the CCAT-R using three videotaped observations and then, in conjunction with the institute, designed a study to measure changes in the quality of caregiver-child interactions before and after the caregivers participated in the program. The findings were mixed, but encouraging. While evaluators found little change in levels of caregiver nurturing, as well as ambivalence among non-parental caregivers about the child-rearing practices of the parents,⁵⁶ scores for engagement increased, with caregivers scoring in the “good” range up from “acceptable.” Ratings increased for back-and-forth communication between the youngest children and their caregivers, which means that “the adults were talking to the children, engaged in activities with them and/or holding them, and the children were engaged with materials more than half the time.”⁵⁷

and child care programs are often evaluated to earn additional stamps of approval. Preschools aiming for NAEYC accreditation undergo third-party observations of one hour per accreditation visit in multiple classrooms covering all ages. Some programs must go through a Program Quality Assessment (PQA) from the HighScope Educational Research Foundation that includes an hour-long observation with raters evaluating multiple measures of quality on a scale of 1 to 5. Teachers in Head Start, the federal preschool program for children in poverty, are accustomed to formal evaluations: Federal monitors arrive in Head Start centers every three years to review accounting books, personnel procedures, and the quality of classroom environments. Officials are now proposing that these observations be used to make decisions on which Head Start programs must compete for renewal of their grants. (See box below)

State Quality Rating and Improvement Systems (QRIS) are already providing lessons. Those who oversee these observations stress that states need to establish a financial structure to sustain the systems. Administrators must manage observations of multiple classrooms in hundreds of programs, enter and track the results from all those observations, and efficiently communicate with each program about what was observed. Observers must be recruited and trained.

Those who teach the observers must also be trained. For both trainers and trainers-of-trainers, states need to provide refresher courses on how to conduct the evaluations and administer repeated tests of raters' reliability in order to avoid "drift"—a phenomenon in which observ-

ers, after using a tool for several months, start to assign teachers scores that are higher or lower than they should be. "Imagine you have a person who comes out to your program who tends to score higher, versus someone who tends to score lower," said Bridget Hamre, a co-developer of the CLASS. "That one point could be a huge and meaningful difference." One training session is "not sufficient," warns a recent policy brief on best practices from the federal Office of Program Evaluation and Research.⁵⁸

In PreK-3rd Reform

As in pre-K and child care settings, there is little sign that individual teachers in the PreK-3rd grades are being evaluated for hiring or firing based on results from valid, reliable observation measures. While PreK-3rd reform efforts—such as FirstSchool, the New Schools Project in Chicago and demonstration sites in Hawaii—are using teacher-observation instruments to guide professional development, administrators have not broached the possibility of using them for formal evaluation of teachers. At the program level, high-stakes evaluations are virtually non-existent, partly because of the disjointed structure of early education. Although the pre-K part of PreK-3rd reform efforts may be assessed under a state's QRIS, the K-3 part falls under the aegis of public schools, which are not part of QRIS and do not typically invite professionals from outside of the school system to conduct periodic observations.

Even if PreK-3rd programs did invite third-party observers to watch teachers work, many questions remain about how the results from those observations could be used to make personnel decisions. Martha Zaslow, Director of the

Proposed Rules Would Lead to More High-Stakes Observations in Head Start

In 2007, with the reauthorization of the Head Start Act, Congress mandated that Head Start programs be reviewed using "a valid and reliable research-based observational instrument" that assesses multiple dimensions of teacher-child interactions.⁵⁹ When the Obama Administration arrived, the Office of Head Start named the CLASS as its instrument of choice, and in 2009 teachers and directors began to be trained on how the tool works. A year later, as part of an effort to boost quality in Head Start programs, federal officials released a draft of regulations that would identify "low-performing" Head Start centers and require them to compete for renewal of their grants against other non-profit providers or school districts who want to run Head Start centers. Centers would be identified as "low-performing" if they received a score of 1 or 2 (out of 7) on one or more domains of the CLASS among other measures.⁶⁰ Comments on this proposal flooded into the Office of Head Start, and as of this printing, final regulations had not yet been released. If the rules are finalized as written, high-stakes evaluations using CLASS scores will become an integral part of Head Start.

Office for Policy and Communications at the Society for Research in Child Development, and co-editor of *Quality Measurement in Early Childhood Settings*, cautions that most observation tools in early education were not designed to be used for high-stakes purposes. “We don’t have a clear consensus on the level or rating that would provide a clear dividing line between adequate and inadequate performance,” Zaslow said.

In only a few cases have school districts taken the step of including observational data collected not only by principals but by professionals who visit teachers’ classrooms several times a year.

And yet as states develop new systems for evaluating teachers in public schools, observation-based assessments are often named as one of many multiple measures that should be part of a teacher’s portfolio. Policymakers typically envision principals as the prime observers, paying little attention to the tools or rubrics principals may be using, not to mention their consistency and objectivity. In only a few cases have school districts taken the step of including observational data collected not only by principals but by professionals who visit teachers’ classrooms several times a year. One high-profile case is the District of Columbia. The public schools in Washington, D.C. are in the second year of using the IMPACT evaluation system, which includes five observations each year, three of which are conducted by “master educators” who have been trained to use the district’s observation tool.⁶¹ In 2011, the school district adapted its observation system for pre-K and kindergarten teachers to take into account “center time,” “morning meetings” and other parts of the daily routine that are not typical in later grades.⁶² ■

The Potential to Enhance Teachers’ Effectiveness Throughout PreK-12 Education

Observation-based assessments of teaching are not the only way to identify, promote, and reward good teaching. The evidence of their power to help teachers improve, however, demands serious consideration among education policymakers. The use of observation-based tools in early edu-

cation—especially in professional development programs that feature coaching or mentoring provides promising lessons for increasing the effectiveness of teachers across the PreK-12 spectrum.

These tools also have the potential to create some common ground among two seemingly distant camps, with education reformers on one side striving for more accountability among teachers, and educators on the other, wary of penalties applied to teachers whose students do not meet benchmarks of performance.

Too often, heated debates about teacher effectiveness rely on thin evidence about teacher performance and its impact on students. Evidence-based tools that capture teacher-student interaction can help to deepen the discourse, presenting new measures of teacher effectiveness that go beyond student test scores. Regardless of their position on evaluation systems, most education policy experts agree on the need for protocols that do not rely on student test scores alone. A recent paper issued by the Economic Policy Institute states that “although standardized test scores of students are one piece of information for school leaders to use to make judgments about teacher effectiveness, such scores should be only a part of an overall comprehensive evaluation.”⁶³ A Gates Foundation paper describing its large-scale Methods of Effective Teaching project makes a similar argument, recognizing that evaluation systems that use test scores alone cannot provide a full picture of what makes a good teacher and should not be “the exclusive proxy for effectiveness.” As the paper states, “They rarely take into account the full range of what teachers do or the context in which they teach.”⁶⁴

Most important, the use of observation tools can catalyze better teaching. Among practicing teachers, training that is directly connected to a teacher’s specific challenges in a specific classroom can spur an intrinsic desire to improve. Among “pre-service” programs for prospective teachers, these tools offer promise too. Education schools and other preparation programs have been criticized for their inability to produce teachers with the knowledge and practical experience to be effective in the classroom. Concerns about teachers’ training in child and adolescent development have also surfaced. In 2010, the National Council for Accreditation of Teacher Education released a landmark study, *The Road Less Traveled: How the Developmental Sciences Can Prepare Educators to Improve Student Achievement*,

which called upon education schools to do a better job of promoting understanding of how children develop and preparing teachers to build strong relationships with their students to improve school performance. One way to show teachers how those relationships are formed, and how to change their instruction to enhance those relationships, is to provide them with specific examples from their own teaching or the teaching practices of others.

Observation tools are already the subject of new research endeavors in the later grades. The Methods of Effective Teaching (MET) project is the most notable example.⁶⁵ MET, which focuses on fourth grade and up, is a \$45 million, multi-year initiative guided by experts at seven research universities, three non-profit organizations, and a few for-profit education researchers and consultants. Twenty-one teachers comprise an advisory panel. The project is designed to shine new light on the power and validity of six different observation instruments, the CLASS and the Framework for Teaching among them. One outcome of the MET project that could assist early childhood educators and public school teachers alike is a toolkit that will provide “advice and a process for training raters to make consistent observations of classroom practice.”⁶⁶

The power of observation can also be amplified through the use of video and online communication tools.

The power of observation can also be amplified through the use of video and online communication tools. The MET project is testing the use of TEACHSCAPE REFLECT, a new video-capture system that uses a special camera to gather 360-degree images of action within a classroom. So far the camera has been used to capture more than 20,000 lessons in more than 3,000 classrooms across six states. The video comes with software based on the Framework for Teaching. Such tools could spur innovations in both professional development and evaluation as video technology is combined with coaching that centers on classroom observations. ■

The Goal: Integrating Professional Development with Formal Evaluation

Educators and administrators typically talk about professional development and performance evaluation as if they

were two separate systems. They are, in fact, integrally connected. Professional development may not pack much power if it isn’t connected to assessments of teachers’ performance. Officials at the Office of Head Start, for example, made a point of choosing the CLASS as a tool for both evaluating programs and providing a structure for professional development. Bridget Hamre, co-author of the CLASS, says that the tool has “gained traction” in part because it is seen as a serious instrument that can have a direct impact on the way a teacher’s work is perceived by supervisors. On the flip side, to assess teachers without helping them improve is to ignore the classroom realities that teachers face. Jack McCarthy of AppleTree saw this firsthand when his institute first introduced its Quality Indicators tool to teachers and administrators. “The resounding feedback was, we don’t know what to do with this. What’s the next step?” he said. No matter what the instrument, he said, teachers need to be introduced to new strategies and alternative approaches that can be used to solve their specific problems.

A more connected system would include a continuous feedback loop of observation-based assessments, coaching, implementation of new strategies, and more observation. Moreover, better links between professional development programs and high-stakes evaluations would ensure that common standards of good teaching are at the core of efforts to help teachers improve *as well as* decision-making about rewarding particular programs or teachers.⁶⁷ The evolution of Quality Rating and Improvement Systems (QRIS) represents an attempt to make these links. Moving beyond the “R” for ratings, states are striving to implement concurrent professional development systems that reflect the “I” for improvement.”

The integration of professional development and evaluation is especially important in PreK-3rd reform efforts, where teachers are steeped in different cultures, one ruled by the tenets of early child development, the other by the demands of a public school system increasingly under scrutiny. Forging connections between professional development and evaluation, as well as collaborative training across the PreK-3rd spectrum, can help to overcome feelings of division. When teachers in early childhood programs and at different grade levels are encouraged to “speak the same language” as they talk about improving instruction and evaluation according to similar benchmarks, they may begin to see themselves as in the same boat, working together under the same set of expectations. ■

Conclusion

Teacher effectiveness will continue to be at the core of education reform across the full spectrum of a child's early experiences and later schooling. There is an urgent need to build a sustainable and robust system of professional development and evaluation that provides teachers with objective measures and the individualized feedback they need to become better at their work. Valid and reliable observation tools that focus on teacher-child interactions should become one of the building blocks for this system. Without a focus on improving interactions—without an emphasis on fostering the deep back-and-forth communication about concepts and skills that leads to learning—children will be at risk of experiencing inconsistent and often mediocre instruction, and achievement gaps will remain. Closing those gaps will require policies that do a much better job of identifying, promoting, and rewarding good teaching. ■

Recommendations

To provide children with engaging learning experiences and high-quality instruction across the age spectrum, from birth through third grade and on into their later school years, policymakers should embrace the use of research-based tools for observing teachers and other professionals who work with children. These tools pinpoint how teachers and other professionals can improve their interactions with children to promote their social-emotional and cognitive growth. They are also essential pieces of an evidence-based approach to ensuring that public investments in education are directed to supporting and promoting teachers who are most effective at engaging children in learning and fostering their success.

To move to a new paradigm centered on fair and reliable measurement of teachers' actual practice will require changes in policies across multiple levels of government and among myriad decision-makers. **Regardless of who is making the policy, however, we offer five general guidelines:**

1. Identification of effective teaching in infant-and-toddler care and across the PreK-12 spectrum—whether in teacher-preparation programs, in-service professional development programs, or personnel evaluation systems—should **include results from valid and reliable observations** of teachers interacting with children.

2. Observation tools for assessing good teaching should be **aligned with standards and assessments across children's ages and grade levels** so that teachers and professionals in one setting, such as a pre-kindergarten classroom, are able to “speak the same language” and share values related to high-quality teaching with teachers and professionals in another setting, such as a kindergarten or first-grade classroom.
3. Policymakers and educators in infant-and-toddler care and across the PreK-12 spectrum (including administrators in all settings) should receive **training in the purposes and implications of observation-based assessments** as well as how to interpret the data from those assessments fairly to improve interactions between children and the adults helping them learn.
4. Professional development and high-stakes evaluations of programs and individual teachers should be aligned to **ensure that all teachers' trainings and evaluations are based on common definitions** of effective teaching; if used in high-stakes evaluations, valid and reliable observation tools for assessing teachers should also be at the core of programs to help them improve.
5. Researchers should **continue to develop and improve observation tools for identifying effective teaching**, with attention given to English language learners and the association between specific teaching practices and children's outcomes in different academic subjects and across multiple domains, including social-emotional and cognitive growth.

We also recommend the following steps be taken at various levels of government, among researchers and educators, and within higher education institutions and professional development initiatives.

The federal government should:

1. Fund large-scale research and implementation projects that use teacher observation tools for evaluation and professional development; require those projects to include reports of associations between

particular aspects of teachers' practice and children's outcomes.

2. Encourage states, local school districts, Head Start and other early childhood programs to use valid and reliable teacher observation tools in the development of evaluation and improvement systems.
3. Highlight and reward states, local school districts, Head Start and other early childhood programs that employ those tools to promote alignment of professional standards between early childhood settings and public schools.
4. Highlight and reward states, local school districts, Head Start and other early childhood programs that use those tools to integrate the formal evaluation of teachers and programs with ongoing, effective professional development.
5. Favor strong examples of the use of teacher-observation tools in grants and appropriations to teacher preparation programs and professional development providers across the PreK-12 spectrum.

States should:

6. Ensure that new designs for teacher-evaluation systems include the use of valid and reliable observational tools that focus on teacher-student interactions and that employ well-trained, third-party professionals to conduct the observations.
7. Continue to develop and refine Quality Rating and Improvement Systems (QRIS) so that infant and toddler programs are included and so that at least one of the multiple measures for rating early childhood programs includes the use of valid and reliable observation tools that focus on adult-child interactions and the use of well-trained, third-party professionals to conduct the observations.
8. Establish guidelines for using observation tools in accordance with research-based examples of best practice. Use of observation tools in Quality Rating and Improvement Systems, for example, should be employed as outlined in the 2011 pol-

icy brief from the Office on Policy Research and Evaluation⁶⁸ to ensure that raters are trained professionally and provided with periodic re-training to ensure that they remain consistent in how they score teachers and programs.

9. Dedicate funds for the development and sustainability of comprehensive assessment and improvement systems that use valid and reliable observation tools, including funding for training observers, providing meaningful reports on progress in early childhood programs and schools, and enabling caregivers and teachers to receive ongoing support and technical assistance using these tools.

Local educators and leaders (including school district leaders, principals, and directors of early childhood programs) should:

10. Participate in training on the use of observation tools to gain a greater understanding of the types of interactions and classroom strategies that foster learning.
11. Give teachers and caregivers opportunities to be observed and assessed using valid and reliable tools, provide them with access to assessment results and videos of their practice, and make time for them to work with professionals who can provide research-based strategies for improvement using those assessments.

Teacher-preparation programs should:

12. Customize coursework and clinical experiences for prospective teachers to familiarize them with observation tools, teaching strategies that enhance teacher-child interactions, and the importance of reflecting on one's own teaching.
13. Require prospective teachers to be observed and assessed using valid and reliable tools, provide those teachers with access to assessment results and videos of their practice, and provide research-based strategies for improvement using those assessments.

Developers of professional development initiatives should:

14. Use valid and reliable observation tools to customize interventions for teachers; develop models that provide teachers with ongoing, individualized support and feedback based on assessment with observation tools.

Researchers should:

15. Study the validity and reliability of existing observation tools in multiple settings and with children of varying backgrounds, including English language learners.

16. Continue to develop observation tools based on the latest findings in science about how children's interactions with adults affect their learning.

17. Expand research to include teacher aides, assistant teachers, directors, principals, and other administrators in the public schools, as well as professionals in multiple types of child care settings, to identify what is needed to improve adults' interactions with children. 

Interviews Conducted

Rebecca Berlin, Client Services Director, Teachstone

Allison Boothe, Tulane, Clinical Assistant Professor, Tulane University School of Medicine, Department of Psychiatry and Neurology.

Janet Burke, Program Director, Child Development Services, ChildSavers, Richmond, VA

Maura Burke, Early Childhood Coordinator, Fairfax County Public Schools, VA

Brenda L. Cubero, Local Coordinator for Family Child Care Homes, Virginia Star Quality Initiative

Charlotte Danielson, Founder, The Danielson Group

Amy Dombro, Early Childhood Consultant

Laura Ensler, Early Childhood Consultant

Ann S. Epstein, Senior Director, Curriculum Development, HighScope Educational Research Foundation

Karen Freel, Vice President, Research & Evaluation, Ounce of Prevention

Linda Gillespie, Technical Assistance Manager, National Infant Toddler Child Care Initiative, ZERO TO THREE

Morgan Green, Director, Virginia Star Quality Initiative

Kim Guieb-Kang, P-3 Project Coordinator, Farrington Complex, Honolulu, HI

Bridget Hamre, Associate Director, Center for Advanced Study of Teaching and Learning, University of Virginia

Phyllis Ida, P-3 Project Coordinator, Windward Demonstration Project, Honolulu, HI

Elizabeth Isakson, Secretary, New York Zero-to-Three Network

Rebecca Klein, Program Developer, Birth to Three Center-Based Training Initiative, Ounce of Prevention

Kerry L. Kriener-Althen, Evaluation Team Director, Center for Child and Family Studies, WestEd

Lourdes Lambert, Assistant Director, Program Implementation, Teachstone

Mary Anne Lesiak, Director of Programs, AppleTree Institute for Education Innovation

Debbie Lund, Center Director, Baby's Space & Tatanka Academy, Minneapolis, MN

Peter Mangione, Co-Director, Center for Child and Family Studies, WestEd

Erin Mason, Data Collection Coordinator, FirstSchool

Chris Maxwell, Director, New Schools Project, Erikson Institute

Jack McCarthy, Managing Director, AppleTree Institute for Education Innovation

Geoffrey Nagle, Director, Tulane University Institute of Infant and Early Childhood Mental Health

Sarah Neville-Morgan, Deputy Executive Director, CA Early Learning Advisory Council

Debra Pacchiano, Director, Continuous Quality Improvement, Ounce of Prevention

Bentley Ponder, Director of Research and Evaluation, Department of Early Care and Learning, Georgia

Sharon Ritchie, Senior Scientist, FPG Child Development Institute, University of North Carolina and Director, FirstSchool

Terrie Rose, Founder and President, Baby's Space, Minneapolis, MN

Schellee Y. Rocher, Director of Provider Operations, Los Angeles Universal Preschool

Jason Sachs, Director, Early Childhood Education, Boston Public Schools

Diana Schaack, Researcher, San Francisco State University

Eva Marie Shivers, Founder and Director, Indigo Cultural Center, AZ

Kathryn Tout, Co-Director, Early Childhood Research & Senior Research Scientist, Child Trends

Gerrit Westervelt, Executive Director, BUILD

Kerrie Welsh, Director, South Carolina Program for Infant/Toddler Care

Christina Weiland, Project Director, Preparing to Succeed Study, Harvard Graduate School of Education and Research Consultant, Boston Public Schools

Karen Yarbrough, Senior Manager, Ounce of Prevention Fund

Notes

¹ Donald J. Hernandez, "Double Jeopardy: How Third-Grade Reading Skills and Poverty Influence High School Graduation," (Annie E. Casey Foundation, April 2011), 3-6.

² For a discussion about school districts that are exceptions, see "Teacher Evaluation, How Observation Fits In," an online forum moderated by Elena Silva and Susan Headen of Education Sector in August 2011. The transcript is available at <http://www.educationsector.org/events/online-discussion-teacher-evaluation-how-observation-fits#cover-it-live>.

³ W. Steven Barnett, "Effectiveness of Early Intervention," *Science* 333 (August 19, 2011): 975-978; Arthur J. Reynolds, Judy A. Temple, Barry A. B. White, Suh-Ruu Ou, and Dylan L. Robertson, "Age 26 Cost-Benefit Analysis of the Child-Parent Center Early Education Program," *Child Development* 82 (January/February 2011): 379-404; Robert C. Pianta, W. Steven Barnett, Margaret Burchinal, and Kathy R. Thornburg, "The Effects of Preschool Education: What We Know, How Public Policy Is or Is Not Aligned with the Evidence Base, and What We Need to Know," *Psychological Science in the Public Interest* 10 (2009): 49-88; W. Steven Barnett, "Preschool Education and Its Lasting Effects: Research and Policy Implications," (Boulder and Tempe: Education and the Public Interest Center & Education Policy Research Unit, 2008); Bridget K. Hamre and Robert C. Pianta, "Can Instructional and Emotional Support in the First-Grade Classroom Make a Difference for Children at Risk of School Failure?" *Child Development* 76 (September/October 2005): 949-967.

⁴ Karen M. LaParo, Bridget K. Hamre, J. Locasale-Crouch, Robert C. Pianta et al, "Quality in Kindergarten Classrooms: Observational Evidence for the Need to Increase Children's Learning Opportunities in Early Education Classrooms," *Early Education and Development* 20 (2009): 657-692; Andrew Mashburn, Robert C. Pianta, Bridget K. Hamre, Jason T. Downer, Oscar Barbarin, Donna Bryant, Margaret Burchinal, Richard Clifford, Diane Early, and Carolee Howes, "Measures of Classroom Quality in Pre-Kindergarten and Children's Development of Academic, Language and Social Skills," *Child Development* 79 (2008): 732-749.

⁵ Diane Marie Early, Oscar Barbarin, Donna M. Bryant, Margaret Burchinal, Florence Chang, Richard M. Clifford, Gisele M. Crawford, Carolee Howes, Sharon Ritchie, Marcia

E. Kraft-Sayre, Robert C. Pianta, W. Steven Barnett, and Wanda Weaver, "Pre-Kindergarten in Eleven States: NCEDL's Multi-State Study of Pre-Kindergarten & Study of State-Wide Early Education Programs (SWEEP): Preliminary Descriptive Report," (National Center for Early Development & Learning, May 24, 2005).

⁶ Robert C. Pianta, Jay Belsky, Renate Houts, and Fred Morrison, "Opportunities to Learn in America's Elementary Classrooms," *Science* 315 (March 20, 2007): 1795-96.

⁷ Stephen B. Plank and Barbara Condiffe, "Pressures of the Season: A Descriptive Look at Classroom Quality in Second and Third-Grade Classrooms," (Baltimore Education Research Consortium, February 2011).

⁸ Robert C. Pianta and Bridget K. Hamre, "Conceptualization, Measurement, and Improvement of Classroom Processes: Standardized Observation Can Leverage Capacity," *Educational Researcher* 38 (March 2009): 115.

⁹ Ibid.

¹⁰ Tamara Halle, Nicole Forry, Elizabeth Hair, Kate Perper, Laura Wandner, Julia Wessel, and Jessica Vick, "Disparities in Early Learning and Development: Lessons from the Early Childhood Longitudinal Study – Birth Cohort (ECLS-B)," (Washington, DC: Council of Chief State School Officers and Child Trends, June 2009).

¹¹ NICHD Early Child Care Research Network, "Child-Care Structure-Process-Outcome: Direct and Indirect Effects of Child-Care Quality on Young Children's Development," *Psychological Science* 13 (2006): 199-206.

¹² *State of America's Children* (Washington, D.C.: Children's Defense Fund, 2010): G-21.

¹³ U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, *National Assessment of Educational Progress* (NAEP), 2011. (To view fourth grade results by family income level, race and ethnicity, go to http://nationsreportcard.gov/reading_2011/nat_g4.asp and click on achievement levels.)

¹⁴ For more on this topic in K-12 education see *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness* (The New Teacher Project,

2009). For a study of this disconnect in pre-K programs, see Diane Early, Kelly L. Maxwell, Margaret Burchinal, et al. "Teachers' Education, Classroom Quality, and Young Children's Academic Skills: Results from Seven Studies of Preschool Programs," *Child Development* 78 (March/April 2007): 558-580.

15 Several organizations and the news media have written about these evaluation-system reforms. For an education-reform-focused analysis of how states are changing their policies, see "Recent Teacher Effectiveness Legislation: How Do the States Stack Up?" (Bellwether Education Partners, August 2011). For arguments against the emphasis on test scores in these systems, see Linda Darling-Hammond, "Recognizing and Developing Effective Teaching: What Policy Makers Should Know and Do," (Partnership for Teacher Quality, May 2010): 8.

16 For more on pre-K-3rd reforms, see Paul Nyhan, "The Power of PreK-3rd: How A Small Foundation Helped Push Washington State to the Forefront of the PreK-3rd Movement," (Foundation for Child Development, October 2011); Lisa Guernsey and Sara Mead, "A Next Social Contract for the Primary Years of Education," (New America Foundation, March 2010); Geoff Marietta, "Lessons for PreK-3rd from Montgomery County Public Schools," (Foundation for Child Development, December 2010); and Sara Mead, "Education Reform Starts Early: Lessons from New Jersey's PreK-3rd Reform Efforts," (New America Foundation, December 2009).

17 Tamara Halle, Jessica Vick Whittaker, and Rachel Anderson, "Quality in Early Childhood Settings: A Compendium of Measures, Second Edition," *Child Trends*, 2010. Prepared by Child Trends for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services; See also Martha Zaslow, Ivelisse Martinez-Beck, Kathryn Tout, and Tamara Halle, eds., *Quality Measurement in Early Childhood Settings*, (Baltimore: Paul H. Brookes Publishing, 2011).

18 Robert C. Pianta, Karen M. LaParo, and Bridget K. Hamre, *Classroom Assessment Scoring System Manual: Pre-K* (Baltimore: Paul H. Brookes Publishing, 2008).

19 Margaret Burchinal et al, "Predicting Child Outcomes at the End of Kindergarten from the Quality of Pre-Kindergarten Teacher-Child Interactions and Instruction," *Applied*

Developmental Science 12 (2008); Robert C. Pianta, "Making Early Childhood Education Matter," *Education Next* (Winter 2007): 47.

20 Burchinal et al, *Applied Developmental Science* (2008); Mashburn et al, *Child Development* 79 (2008): 732-749; Carollee Howes, Margaret Burchinal, Robert Pianta, Donna Bryant, Diane Early, Richard Clifford, Oscar Barbarin, "Ready to Learn? Children's Pre-Academic Achievement in Pre-Kindergarten Programs," *Early Childhood Research Quarterly* 23 (2008): 27-50; see also "How Does the CLASS Relate to Students' Academic and Social Development (predictive validity)?" *Classroom Assessment Scoring System Manual: Pre-K*, (2008): 104.

21 Margaret Burchinal, Kristen Kainz, and Yaping Cai, "How Well Do our Measures of Quality Predict Child Outcomes? A Meta-Analysis and Coordinated Analysis of Data from Large-Scale Studies of Early Childhood Settings," in Zaslow et al, eds., *Quality Measurement in Early Childhood Settings*, 2011.

22 *Statistical Abstract of the United States* 2012, "Table 236: Children Who Speak a Language Other Than English at Home, By Region: 2009," U.S. Census Bureau.

23 Email correspondence with Gerrit Westerveld, October 27, 2011.

24 Andrew Mashburn and Robert C. Pianta, "Opportunity in Early Education: Improving Teacher-Child Interactions and Child Outcomes," in Arthur Reynolds et al, eds, *Childhood Programs and Practices in the First Decade of Life: A Human Capital Integration* (Cambridge, Mass: Cambridge University Press, 2010); Robert C. Pianta and Bridget K. Hamre, "Classroom Processes and Positive Youth Development: Conceptualizing, Measuring, and Improving the Capacity of Interactions Between Teachers and Students," *New Directions for Youth Development* 121 (Spring 2009): 33-46; and J.P. Allen, A.Y. Mikami, and R.C. Pianta, "An Interaction-Based Approach to Enhancing Secondary School Instruction and Student Achievement," *Science* 333 (August 19, 2011): 1034-1037.

25 Heather Sandstrom, Shannon Moodie, and Tamara Halle, "Beyond Classroom-Based Measures for Preschoolers: Addressing the Gaps in Measures for Home-Based Care and Care for Infants and Toddlers," in *Quality Measurement in Early Childhood Settings*, eds, Martha Zaslow, Ivelisse

Martinez-Beck, Kathryn Tout and Tamara Halle (Baltimore: Brookes Publishing Co, 2011).

26 West Ed, "A New Yardstick for Assessing Early Childhood Care," *R&D Alert*, 8(1), 18-21, retrieved June 14, 2011 at http://www.wested.org/online_pubs/rd-06-01.pdf.

27 Kerry Kriener-Althen and Peter M. Mangione (2010), The PITC Program Assessment Rating Scale: Assessing Structural, Global, and Process Quality in Infant-Toddler Care, Presentation 3 in Symposium: "Defining, Providing, and Assessing Quality Child Care in Group Settings that Support Infants and Toddlers," American Educational Research Association Annual Meeting, Denver, CO.

28 J. Ronald Lally and Peter. L. Mangione, "The Program for Infant/Toddler Care," in *Approaches to Early Childhood Education*, eds., J.L. Roopnarine and J.E. Johnson, 2009, Upper Saddle River, NJ: Pearson Education, Inc.

29 Interview with Kerry Kriener-Althen, August 9, 2011.

30 Kriener-Althen and Mangione, 2010.

31 From presentation by Therese Ahlers, director of the Wisconsin Alliance for Infant Mental Health, in "Integrating Social and Emotional Development into Early Childhood Systems: Moving Practice into Policy," ZERO TO THREE National Training Institute, December 9, 2010.

32 Walter S. Gilliam, "Prekindergarteners Left Behind: Expulsion Rates in State Prekindergarten Systems," (A.L. Mailman Family Foundation and Foundation for Child Development: May 4, 2005).

33 Sheryl Scott Heller, Allison Boothe, Angela Keyes, et al. "Implementation of a Mental Health Consultation Model and its Impact on Early Childhood Teachers's Efficacy and Competence," *Infant Mental Health Journal*, 32 (2011): 143-164.

34 Sheryl Scott Heller, Allison B. Boothe, A. Keyes, Geoffrey Nagle, K. Vaughn, M. Sidell, & J. Rice, "It's All About the Relationship: Mental Health Consultation's Impact on Classroom Interactions," *Early Education and Development*, in press.

35 Kathryn Tout, Rebecca Starr, Margaret Soli, Shannon Moodie, Gretchen Kirby, and Kimberly Boller, "Compendium

of Quality Rating Systems and Evaluations," prepared by Child Trends for the Office of Planning, Research and Evaluation, Administration for Children and Families (June 2010): ES1; Note that the number of QRIS around the country may now be as high as 34, according to Lea J.E. Austin, Marcy Whitebook, Maia Connors, and Rory Darrah, "Staff Preparation, Reward and Support: Are Quality Rating and Improvement Systems Addressing all the Key Ingredients Necessary for Change?" Center for the Study of Child Care Employment, 2011.

36 By law, all Head Start programs are required to undergo observations as part of federal monitoring.

37 Interview with Bentley Ponder, May 12, 2011.

38 Robert C. Pianta, "Coaching and Coursework Impacts on Preschool Teachers' Interactions with Children," presentation for biennial meeting of the Society for Research in Child Development (SRCD) in Montreal, Canada, March 2011.

39 Email correspondence with Susan Neuman, August 10, 2011.

40 Tanya S. Wright, "Online Logs: A Tool to Monitor Fidelity of Implementation in Large-Scale Interventions," in *Preparing Teachers for the Early Childhood Classroom: Proven Models and Key Principles*, 2010, eds., Susan B. Neuman and Michael L. Kamil, Baltimore: Paul H. Brookes Publishing Co.

41 Interview with Jack McCarthy, August 1, 2011.

42 Amanda Williford and Andrew Mashburn, "The Impact of Virginia's Star Quality Initiative on Classroom Quality and Child Outcomes," presentation at SRCD biennial meeting in March 2011.

43 The Emerging Academic Snapshot was among the measures of classroom quality used in the 11-state study of pre-K programs by Early et al, 2005.

44 Interview with Jason Sachs, September 19, 2011.

45 Interview with Phyllis Ida, September 9, 2011.

46 Robert C. Pianta, W. Steven Barnett, Margaret Burchinal and Kathy Thornburg, "The Effects of Preschool Education: What We Know, How Public Policy Is or Is Not Aligned With

the Evidence Base, and What We Need to Know,” *Psychological Science in the Public Interest* 10(2) (2009): 49–88.

47 *Taking Stock: Assessing and Improving Early Childhood Learning and Program Quality*, The Report of the National Early Childhood Accountability Task Force convened by the Foundation for Child Development, the Pew Charitable Trusts and the Joyce Foundation (October 2007): 3.

48 OPRE, “Best Practices for Conducting Program Observations as Part of QRIS,” June 2011, 8.

49 Sandstrom et al, 2011; Marcy Whitebook and Dan Bellm, *Roots of Decline: How Government Policy Has De-Educated Teachers of Young Children*, Berkeley: Center for the Study of Child Care Employment, 2006.

50 “Quality Rating and Improvement Systems (QRIS): Inclusion of Infant/Toddler Quality Indicators,” fact sheet from the National Infant & Toddler Care Initiative (September 2010). Accessed on November 1, 2011 at http://nitci.nccic.acf.hhs.gov/resources/qriss_factsheet_final_version.pdf.

51 Rossana Bisceglia, Michal Perlman, Diana Schaack et al, (2009), “Examining the Psychometric Properties of the Infant-Toddler Environment Rating Scale-Revised Edition in a High-Stakes Context,” *Early Childhood Research Quarterly*, 24, 121-132.

52 Virginia is providing training on the Toddler CLASS as part of the state’s QRIS, and Washington and California are running pilot projects that use the Toddler CLASS.

53 The PARS is currently in use in Ohio’s QRIS, has been recommended for program evaluation in California, and is slated for implementation in other states.

54 See the Institute for a Child Care Continuum, Bank Street College of Education, for further discussion of the Child Care Assessment Tool for Relatives (CCAT-R).

55 Toni Porter and Lee Vuong, *Tutu and Me: Assessing the Effects of a Family Interaction Program on Parents and Grandparents*, New York: Bank Street College of Education, Institute for a Child Care Continuum, 2008.

56 Ibid.

57 Ibid.

58 OPRE, “Best Practices for Conducting Program Observations as Part of QRIS,” June 2011.

59 *Improving Head Start for School Readiness Act of 2007*, Section 641a - (c) (2) (F) Accessed on July 22, 2011 at <http://www.gpo.gov/fdsys/pkg/BILLS-110hr1429enr/pdf/BILLS-110hr1429enr.pdf>.

60 “Designation Renewal of Head Start Grantees,” *Federal Register*, 75 FR 57709. Accessed on October 24, 2011 at <http://www.federalregister.gov/articles/2010/09/22/2010-23583/head-start-program>.

61 For insights on IMPACT, see Susan Headden, “Inside IMPACT: D.C.’s Model Teacher Evaluation System,” (Washington, D.C.: Education Sector, 2011).

62 *IMPACT: The District of Columbia Public Schools Effectiveness Assessment System for School-Based Personnel, Group 2a: Early Childhood Teachers*, DCPS, August 2011.

63 Eva L. Baker, Paul E. Barton, Linda Darling-Hammond, Edward Haertel, Helen F. Ladd, Robert L. Linn, Diane Ravitch, Richard Rothstein, Richard J. Shavelson, and Lorrie A. Shepard, “Problems with the Use of Student Test Scores to Evaluate Teachers,” (Economic Policy Institute, August 2010).

64 *Methods of Effective Teaching Project: Working with Teachers to Develop Fair and Reliable Measures of Effective Teaching*, The Bill and Melinda Gates Foundation (June 2010): 3.

65 The Methods of Effective Teaching (MET) project is testing the use of observation tools in the 4th through 8th grades, Algebra I classrooms, 9th grade English, and high school biology.

66 *Methods of Effective Teaching Project*, 11.

67 Carolee Howes, Robert C. Pianta, Donna Bryant, Bridget K. Hamre, Jason Downer and S. Soliday-Hong, “Ensuring Effective Teaching in Early Childhood Education Through Linked Professional Development Systems, Quality Rating Systems and State Competencies: The Role of Research in an Evidence-Driven System,” *A National Center for Research in Early Childhood Education White Paper*, 2008.

68 OPRE, “Best Practices for Conducting Program Observations as Part of QRIS,” June 2011.

MAIN OFFICE
1899 L Street, NW
Suite 400
Washington, DC 20036
Phone 202 986 2700
Fax 202 986 3696

CALIFORNIA OFFICE
921 11th Street
Suite 901
Sacramento, CA 95814
Phone 916 448 5189
Fax 916 448 3724



WWW.NEWAMERICA.NET